Applied A.I. Solutions

Data Visualization Techniques

Professor

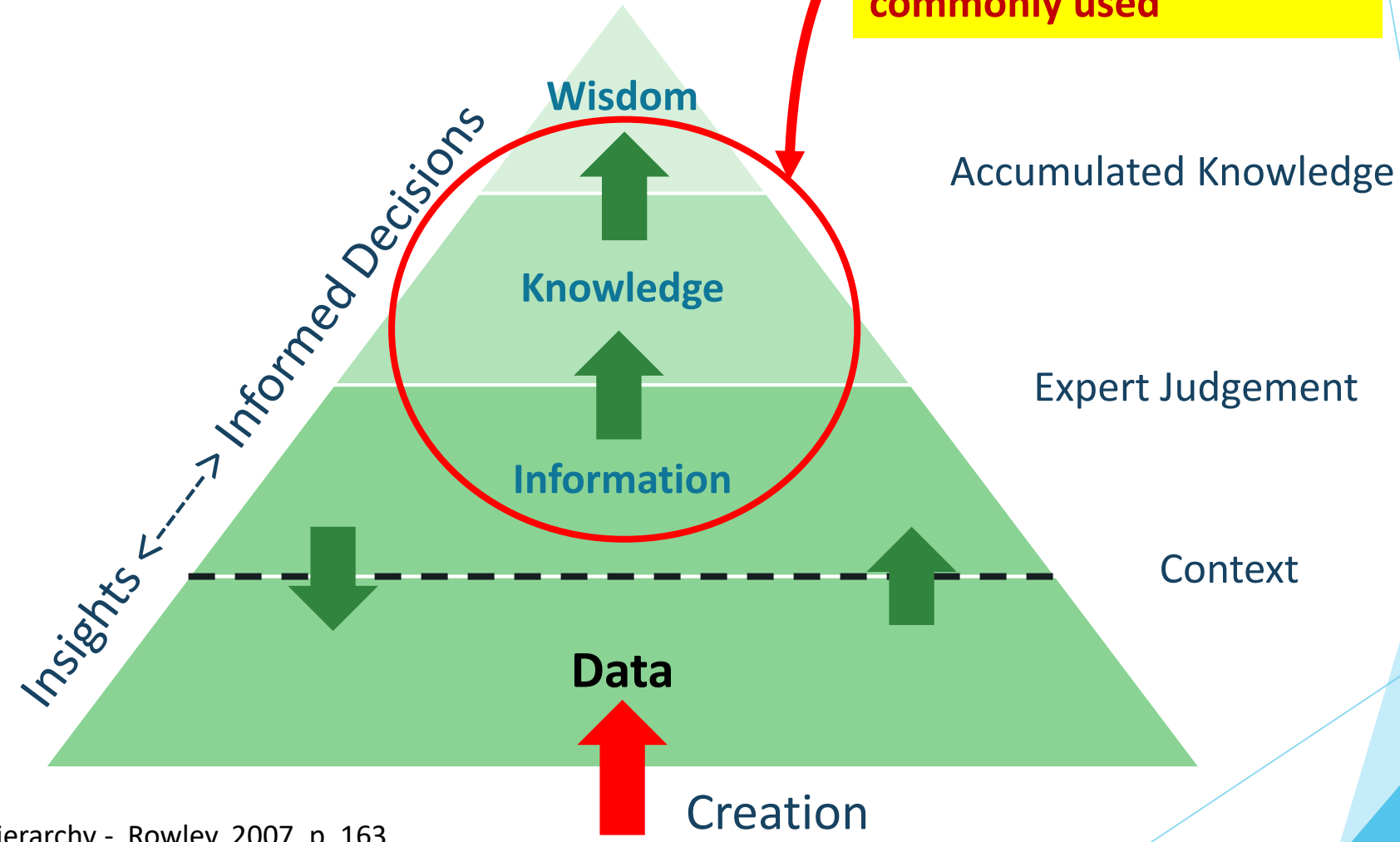
Daniel Vitaver, EMBA

daniel.vitaver@georgebrown.ca

Data Value Extraction, Process Automation, Collaboration, Communication and Research

The DIKW Hierarchy

Where Data Visualizations Techniques and Tools are commonly used



The DIKW Hierarchy - Rowley, 2007, p. 163

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Visualization Techniques

Data Visualization Techniques and Tools help to

to understand data

to empower analytics

- to uncover hidden patterns and trends
- to see the story that lives within that data

to get relevant insights

- for informed decision-making
- for performance management

Data Visualization Techniques – Essential Concepts

- The **value** of data
- The importance of **context** and **storytelling**
- Introduction to **data types** (structured and unstructured)
- **Process Automation**, Collaboration, Communication, Research and Decision Making

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Visualization Techniques – Data Value Extraction

- Decision-making
- Performance Management
- Innovation – R&D
- Risk Management
- Business Development
- Enterprise Asset Valuation

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Costs of Data

- Collecting, storing data
- Data loss impact
- Risks
- Data improvement

Benefits of Quality Data

- Higher quality
- What data could be sold for
- What competitor would pay for it
- Expected revenue from innovative uses of it

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data types and main characteristics

A data type is an **attribute of data** which tells the compiler or interpreter how the programmer intends to use the data.

Common data types:

- Integer numbers (whole numbers)
- Floating-point or real numbers (contains decimal points)
- Characters (single letters)
- Strings (combination of letters, characters and symbols)
- Booleans (True/False, yes/no)

- constrains
- defines the operations
- set of values

10

Data types	
Uninterpreted	Bit, Byte, Word, Bit array
Numeric	Complex, Decimal, Fixed point, Floating point, Double precision, Extended precision, Long double, Integer, signedness, Interval, Rational
Pointer	Address, physical, virtual, Reference
Text	Character, String
Composite	Algebraic data type, generalized, Array, Class, Dependent, Equality, Inductive, Intersection, List, Object, metaobject, Option type, Product, Record, Set, Union
Logical	Boolean, Collection, Enumerated type, Exception, Function type, Recursive data type, Semaphore, Stream, Type class, Unit type
Object-oriented	Abstract data type, Data structure, Generic, Kind, metaclass, Object type, Parametric polymorphism, Primitive data type, Protocol, interface, Subtyping, Type constructor, Type conversion, Type system, Type theory, Variable

[1] Source: Wikipedia, Data Types

DBMS Classification (design or use)

Data Type

Structured
Semi-structured
Unstructured

DB Type

Relational (SQL)
Object-oriented
NoSQL

Model

Relational
Object-oriented
Star schema / Multidimensional

Content Type

Real-time
Distributed
Data Warehouse

Processing Type

OLTP
OLAP
Batch
Real-time
Parallel
In-memory

Data Types

Structured Data

- Structured data refers to all types of data that are **organized** in a any defined form
- Structured data is **defined types of data** in a structure
- Structured data lives in **rows and columns** and it can be **mapped into pre-defined fields**.
- Structured data has a **pre-defined data model**, typically a **relational** data model (SQL database)

Data Types – cont'd

Unstructured Data

- Data that either **does not have a pre-defined data model** or is **not organized in any pre-defined format**. Usually, it is stored in its native format.

- It contains text, videos, audio, images, it may also contain dates, numbers, facts

- Common Techniques & Methodologies for **structuring text** are:

  - Natural Language Processing (NLP)
  - Text Analytics
  - Semantic Tagging with metadata

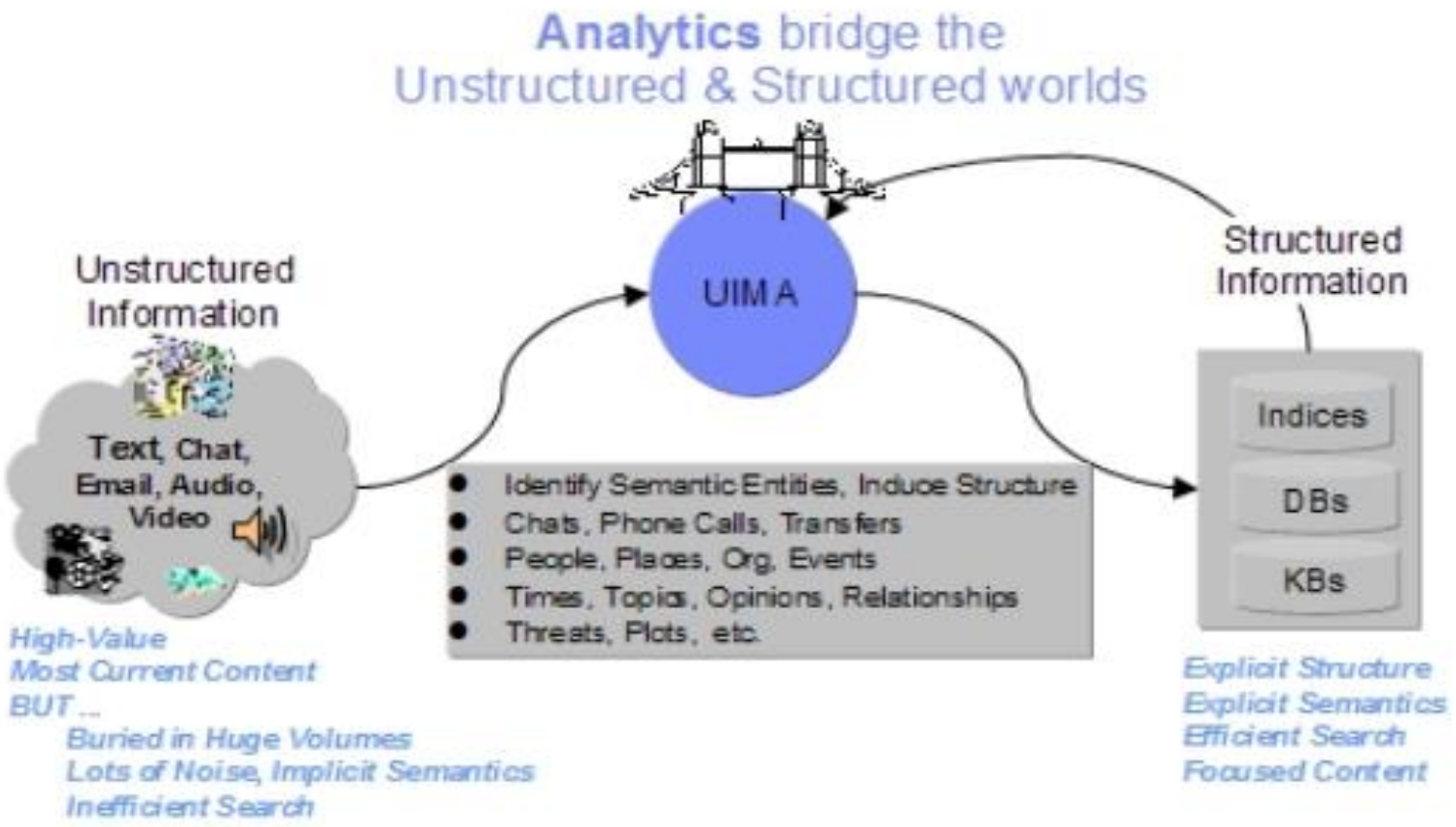
Data Types – cont'd

Unstructured Data

- The **Unstructured Information Management Architecture** (UIMA) standard provides a c**ommon framework for processing this information to extract meaning and create structured data** about the information

- **References**
  - https://uima.apache.org/uima-specification.html
  - https://www.oasis-open.org/

Apache UIMA™ project

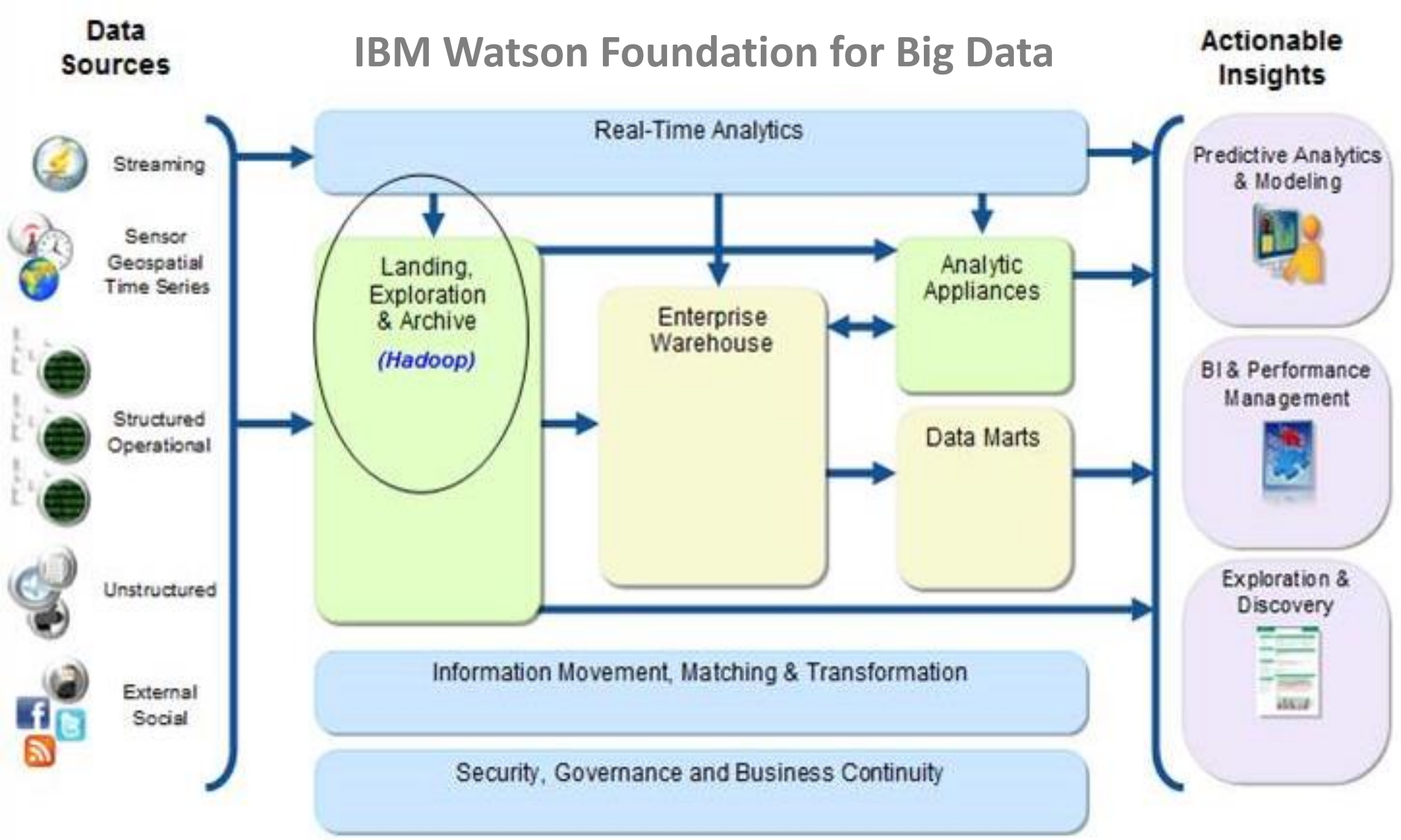


Data Types – cont'd

Semi-structured Data

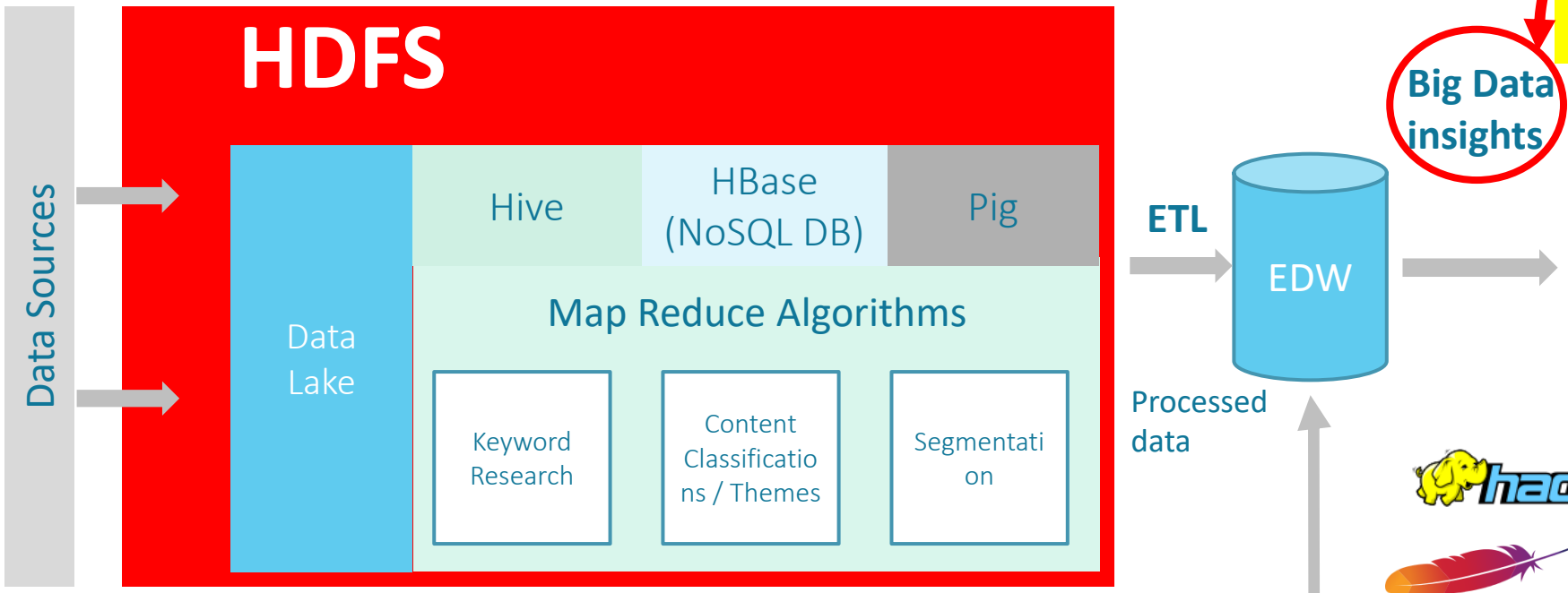
- Mix between structured and unstructured data
- Data has some consistent characteristics but doesn't conform to a rigid structure expected with a relational database
- Some organizational properties such as semantic tags or metadata is used to make it easier to organize

IBM Watson Foundation for Big Data



Big Data Architecture – Hadoop framework + ecosystem

Data load
using Sqoop



Data storytelling

Data storytelling is a structured approach for communicating data insights, and it involves a combination of three key elements: [1]

1. Data (explore)
2. Visuals (assessment and selection)
3. Narrative (explain)

It is an essential Data Science skill everyone needs.

[1] **Brent Dykes, Contributor, Forbes**

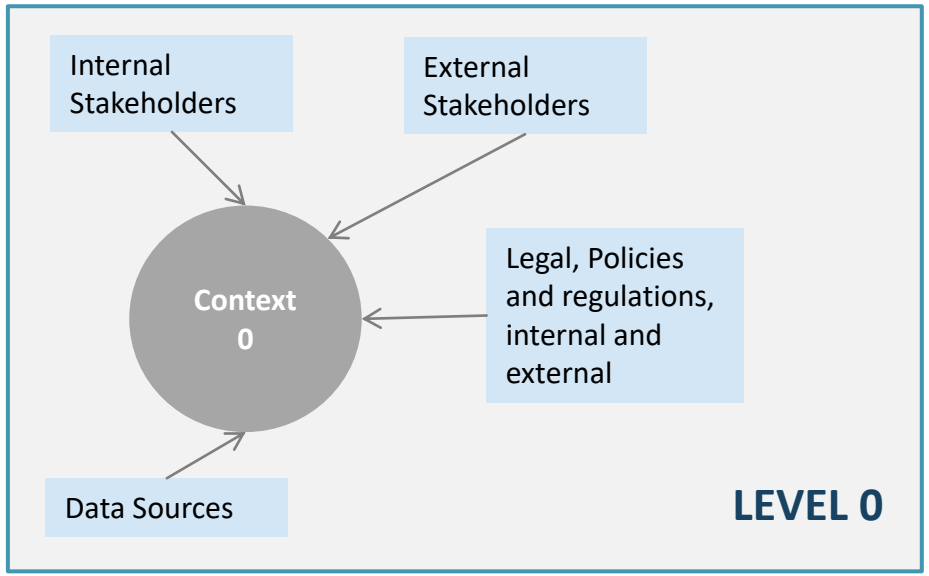
Principles that guide Data Visualization Techniques practice:

- Identify and understand your stakeholders and **audience**
- Set your **goals**
- Define the message and **story** that you want to tell
- Select the **right type of charts**
- Use visualization **toolkit** (shapes, size, formats, colours, layout)
- Set layout **hierarchies and prioritizations**
- Include **comparisons** for analysis
- Tell your story; use **storytelling best practices**

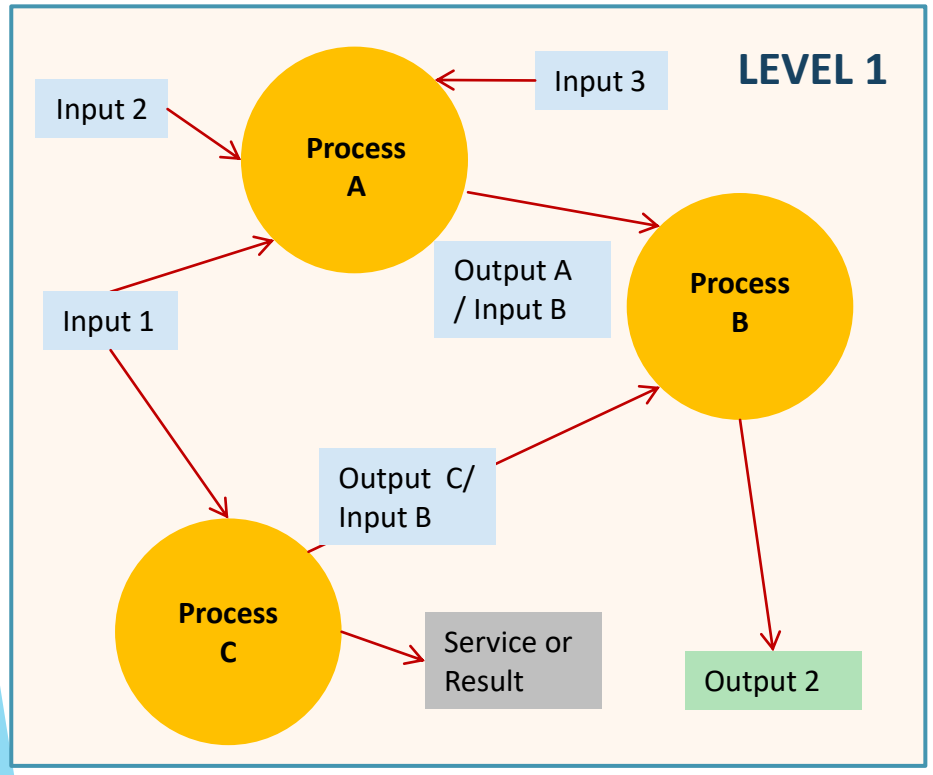
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

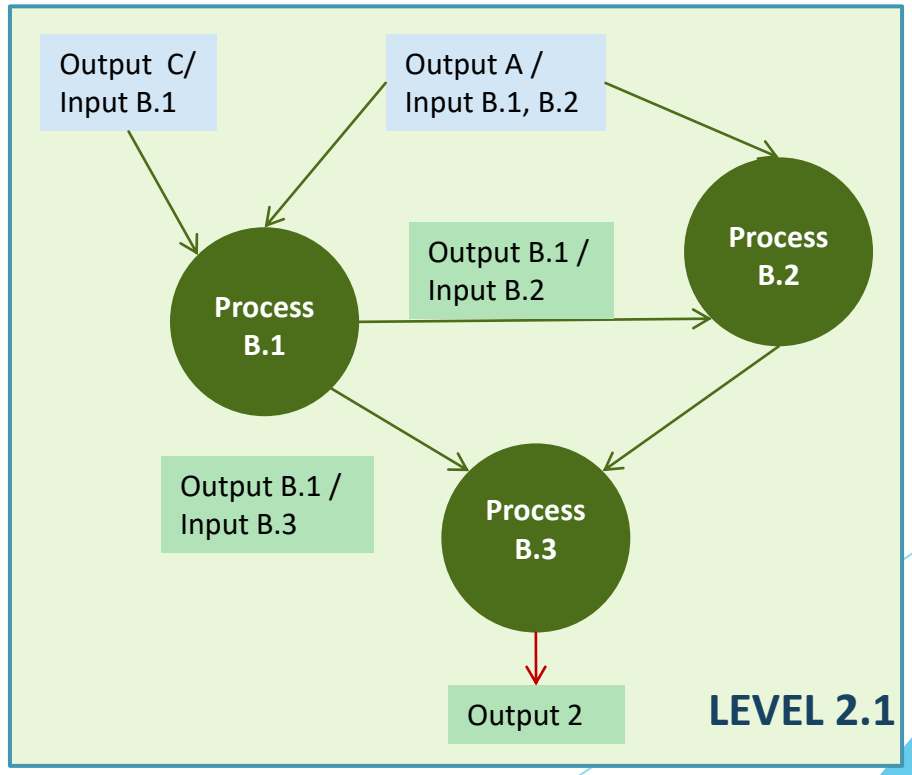
Business Processes

- A process is a set of interrelated **actions** and **activities** performed to achieve a pre-specified product, result, or service
- Each process is characterized by its **input**, the **tools and techniques** that can be applied, and the resulting **outputs**
- Each process can be broken down into other processes, and so on, until we reach a **level** in which simple activities are defined using measurable inputs and measurable outputs


Internal Stakeholders
External Stakeholders
Context 0
Legal, Policies and regulations, internal and external
Data Sources
LEVEL 0

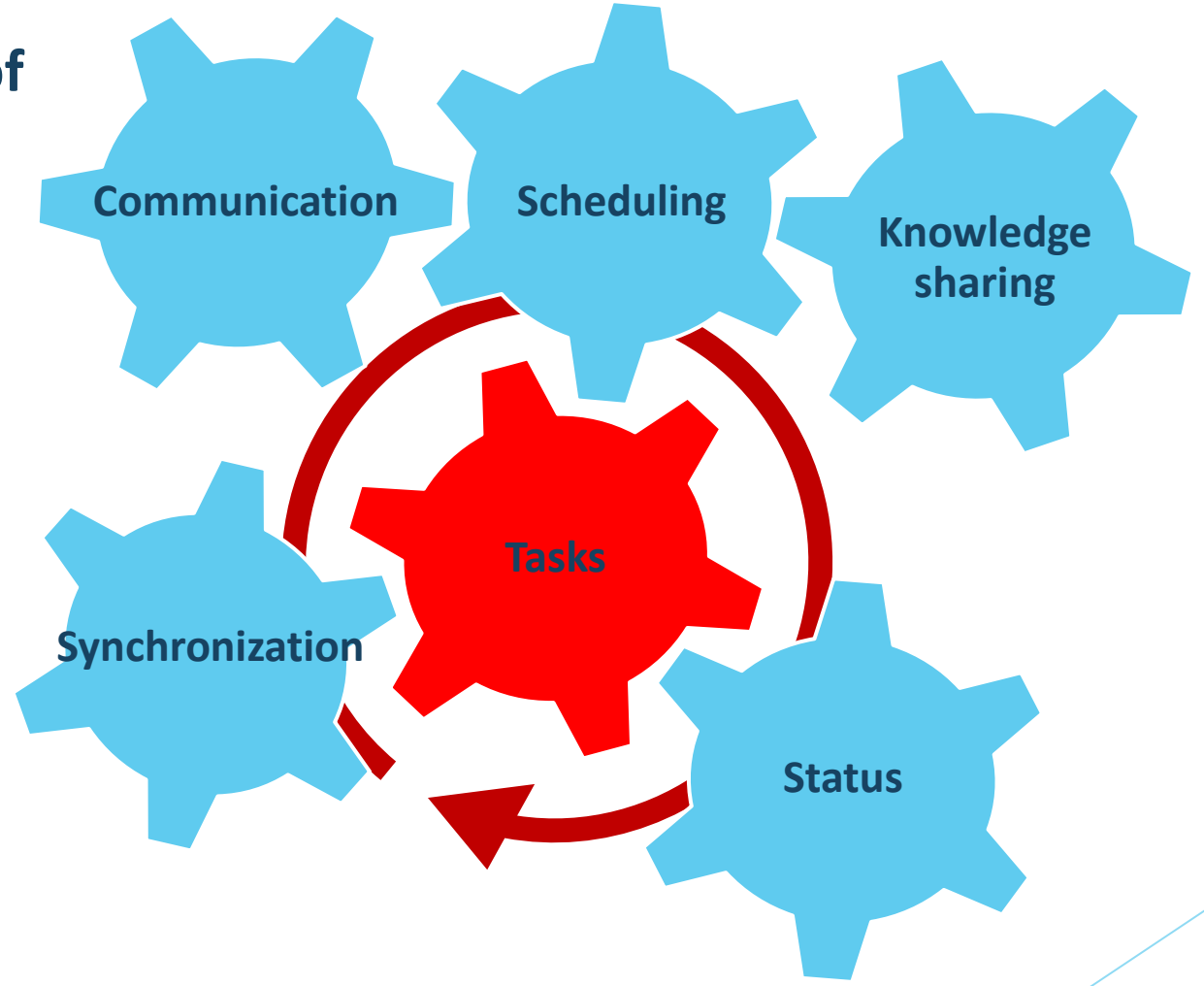

LEVEL 3


LEVEL 1
Input 2
Input 3
Process A
Process B
Output A / Input B
Input 1
Output C/ Input B
Process C
Service or Result
Output 2


Output C/ Input B.1
Output A / Input B.1, B.2
Process B.1
Output B.1 / Input B.2
Process B.2
Output B.1 / Input B.3
Process B.3
Output 2
LEVEL 2.1

Process Automation, Collaboration, Communication

Components of Collaboration



Process Automation Systems

- A process automation system (**PAS**) **automatically control a processes**
- A network is used to interconnect sensors, controllers, operator terminals
- Software and controls regulate equipment to run efficiently, ensuring the consistency of quality, and forecast when maintenance is needed

[1] **Source: Wikipedia**

IT Process Automation (ITPA)

- ITPA **orchestrates** and integrates tools, people, processes through automated workflows

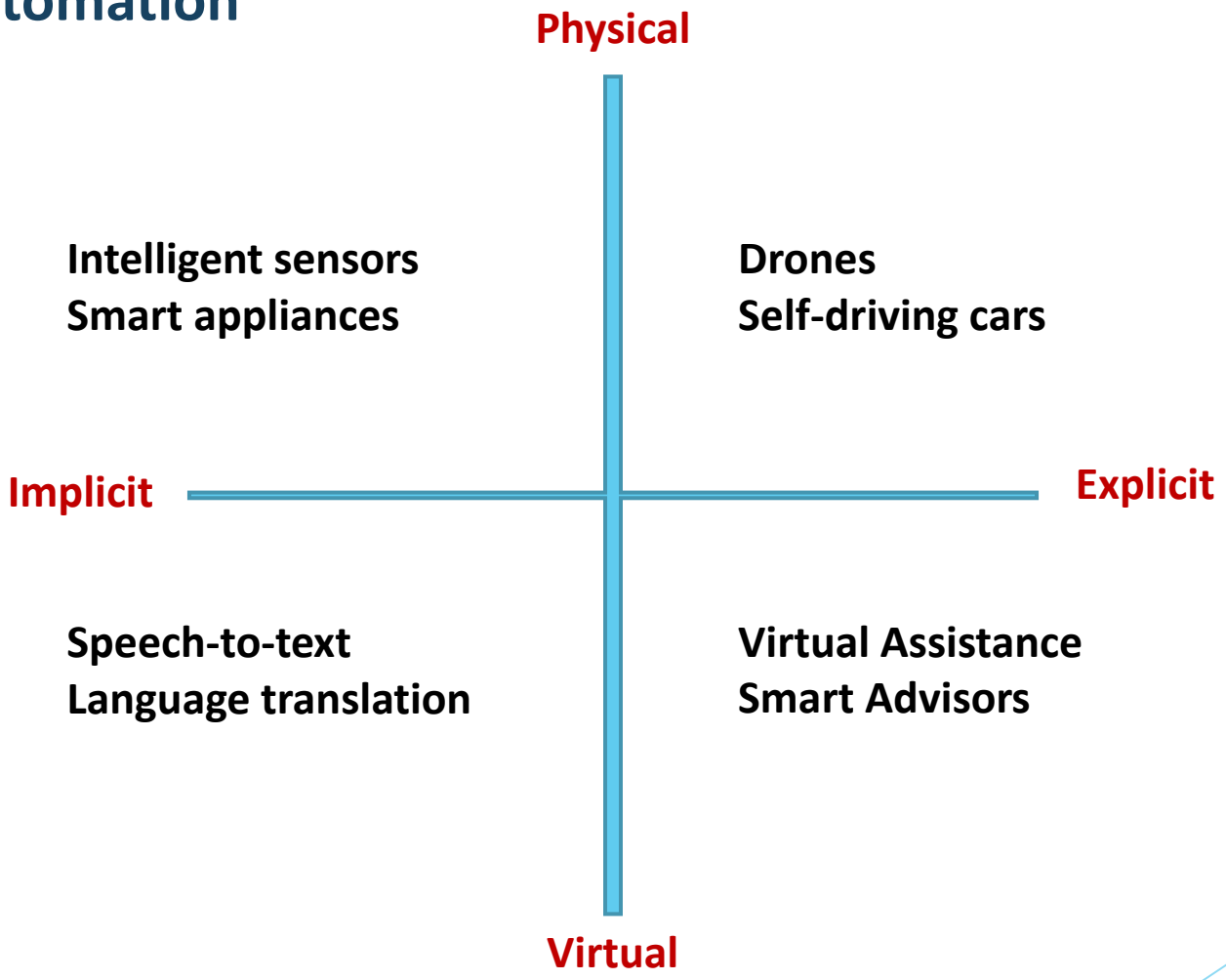
- ITPA uses software applications that are programmed to **perform any repeatable pattern, task or business workflow**

[1] **Source: Wikipedia**

Digital Process Automation (DPA)

- Plain old automation
- Natural language processing
- Machine learning,
- **DPA is not AI/ML**, **but** fosters the use of AI/ML to make it more powerful
- Repetitive tasks done on a predefined cadence or sequence
- Tasks span multiple tools or cross-tool integrations

Applying AI to process automation



Process Automation - Current State

1. Process automation is critical for the success of digital transformation. It's a key component of the IT strategies

2. Complex Processes incorporate many steps and components across different technologies that need be automated

3. **Companies are looking to increase investment in process automation**

4. **Robotics** Process Automation provide relief as a noninvasive form of integration

Reference (in Blackboard)

- Read Camunda's report on "The State of Process Automation 2020"
- Read Gartner's report on "Move Beyond RPA to Deliver Hyperautomation, Dec 2019.

Research and Decision-making

4-type research utilization (Nutley *et al* 2003)

1. **Instrumental use** - research feeds directly into decision-making
2. **Conceptual use** - research changes professionals' understanding providing new ways of thinking
3. **Mobilization of support** - research as an instrument of persuasion
4. **Wider influence** - research findings may come into use through networks of professionals and researchers and alter paradigms, strategies and actions

Reference: https://www.soas.ac.uk/

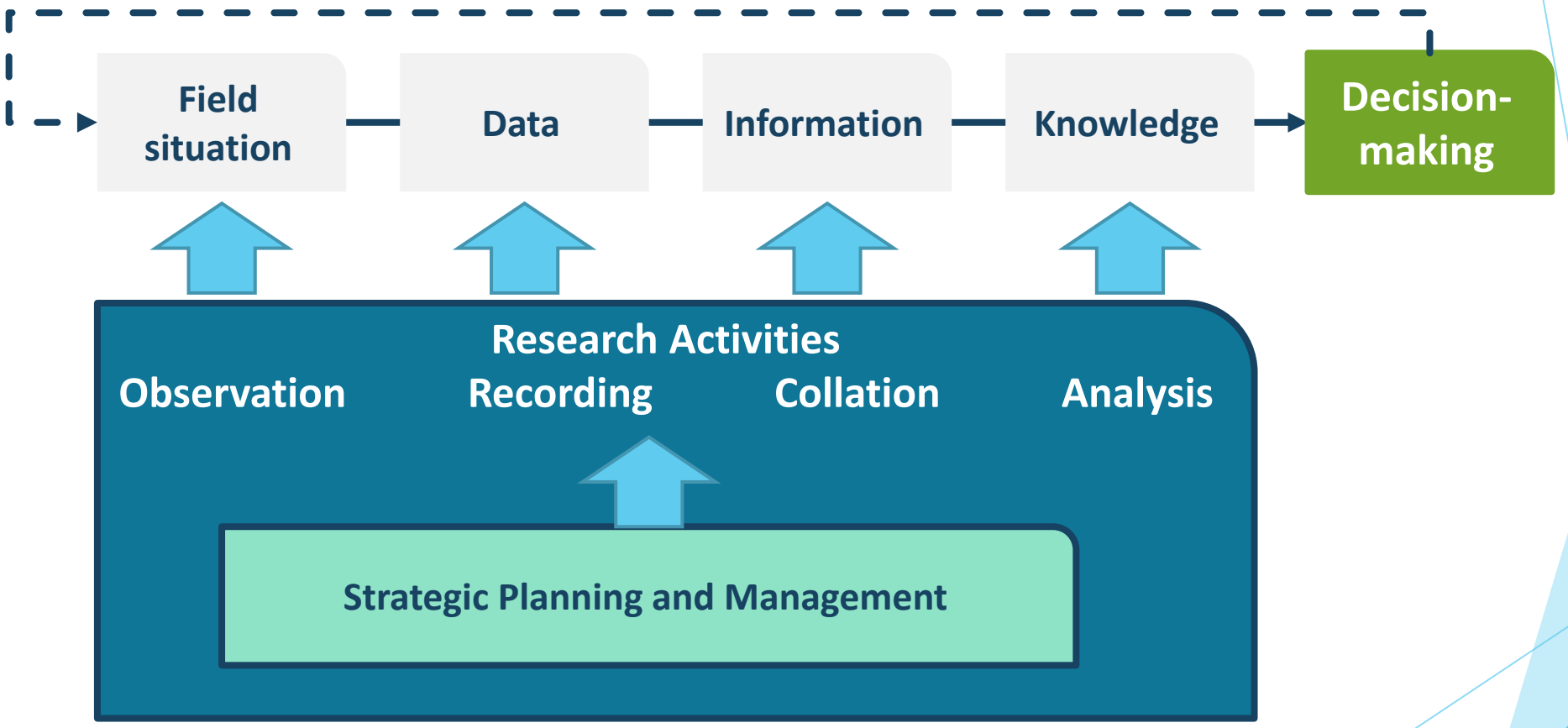
Useful information for decision-making - qualities

- **Comprehensive**
- Consistent with previous studies, regarding methods, **facts and conclusions**
- **Clearly presented**, easy to comprehend
- **Reliable** and **accurate**
- Representative of the situation as a whole, **not biased** or one-sided
- **Timely**
- Directed and delivered to the **relevant decision-makers**
- **Cost-effective**

Information for decision-making - Risks

- **Bias**
- Recording and editing
- Selection (data vs information, lack of metadata)
- Timing
- Analysis and interpretation (selection of inappropriate techniques)

Research, information, and decision-making



Reference
OAS University of London https://www.soas.ac.uk/